Reducing chiral symmetry violations in lattice QCD with domain-wall fermions

Yigal Shamir
School of Physics and Astronomy, Beverly and Raymond Sackler Faculty of Exact Sciences, Tel-Aviv University, Ramat Aviv, 69978 Israel

(Received 8 July 1998; published 5 February 1999)

The inverse of the fermion matrix squared is used to define a transfer matrix for domain-wall fermions. When the domain-wall height M is bigger than one, the transfer matrix is *complex*. Slowly suppressed chiral symmetry violations may then arise from all eigenvalues of the transfer matrix which are located near the unit circle. Using a variable lattice spacing for the fifth coordinate we enforce the strict positivity of the transfer matrix for any M. We furthermore propose a modified pseudo-fermion action, aimed to decrease the density of close-to-unity eigenvalues of the (positive) transfer matrix, at the price of a small renormalization of the coupling constant. We explain why these changes may reduce chiral symmetry violations in lattice QCD simulations. [S0556-2821(99)03705-4]

PACS number(s): 11.15.Ha, 12.38.Gc

I. INTRODUCTION

Numerical QCD simulations [1–4] using domain-wall fermions [5–8] reveal small, but still significant, lattice-artifact violations of chiral symmetries for $N_s \sim 15$, where N_s is the extent of the lattice in the fifth direction. It is important to improve our understanding of such anomalous effects, and, hopefully, to devise new domain-wall actions capable of better suppressing them. Both at the tree level and in perturbation theory the anomalous effects decrease exponentially with increasing N_s. This applies in particular to the additive radiatively induced quark mass [6,8,9,2]. In this paper we will thus focus on non-perturbative anomalous contributions to the chiral Ward identities introduced in Ref. [10] and first studied numerically in Ref. [1].

Free domain-wall fermions have a single chiral zero mode on each four-dimensional boundary of a five-dimensional lattice when the five-dimensional mass term (or ''domain wall height'') M is in the range $0<M<2$. As one moves to the range $2<M<4$ this zero mode disappears, and four new zero modes of the opposite chirality appear on each boundary. Thus, the range $0<M<2$ supports a single (approximately) massless quark, and the range $2<M<4$ supports four (approximately) massless quarks with ''flipped'' chiralities. This pattern generalizes to higher values of M, until for $M>10$ there are no zero modes at all.

With suitable subtractions to ensure its finiteness (to be discussed in detail in this paper) the $N_s \to \infty$ limit of domain-wall fermions can be written down compactly in the overlap formalism [11,12]. In this formalism the (subtracted) fermion partition function is expressed in terms of the overlap of a second-quantized ground state with a reference state or, more generally, as the ground-state expectation value of an operator representing the boundary conditions in the s-direction [10].

In the ''old'' overlap formalism [11], the second-quantized Hamiltonian is (minus) the logarithm of the transfer matrix that hops domain-wall fermions a single site in the s-direction. This formalism has the same massless-quark spectrum as above. The ''new'' overlap formalism [12] involves a different Hamiltonian which arises in a continuous s-coordinate limit. The new overlap (and the related four-dimensional, non-local, chirally invariant action [13]) has the same massless-quark spectrum for $0<M<2$. In the range $2<M<4$ four massless quarks with flipped chiralities appear as before, but the original massless quark *remains* in the spectrum. As M is further increased more massless quarks appear, alongside with all the previous ones that keep staying in the spectrum. This different free-field spectrum agrees with the pattern of level crossing found in a smooth instanton background [14].

A more fundamental difference between the two overlap formulas is found in the properties of the transfer matrix. In the new overlap formalism the transfer matrix is strictly positive for all values of M, simply because it is defined from the outset as $\exp(-H)$, where H is a well-defined local lattice Hamiltonian. In the old overlap formalism, on the other hand, the transfer matrix is positive only for $0<M<1$, as in the framework of domain-wall fermions from which it is derived. We argue below that for $M>1$ both positivity and hermiticity of the domain wall's transfer matrix are lost, and its eigenvalues are in general *complex*.

The aim of this paper is to study what could be the dominant anomalous contributions to chiral Ward identities, and to suggest methods for suppressing them. A key role will be played by a (first quantized) transfer matrix related to the inverse of the fermion matrix squared. For flavor non-singlet chiral symmetries, the anomalous term in the lattice Ward identities [10] involves the correlations of fermion operators on the $s=1$ and $s=N_s$ boundaries with fermion operators at $s \sim N_s/2$. Such correlations decrease exponentially if the transfer matrix has no eigenvalue with an absolute value close to one. Conversely, if there are many such eigenvalues, one expects a much slower (presumably power-law) decrease of anomalous correlations.

Numerical results [1,3] suggest an optimal value of M between 1.6 and 1.7 for $\beta \sim 6$. In this range, the transfer matrix of standard domain-wall fermions is complex. Slowly decreasing anomalous contributions may then arise from all eigenvalues of the transfer matrix which are close to the unit circle.

The positivity of the transfer matrix can be enforced for any M by restricting the range of a_5, the lattice spacing for the fifth coordinate [15], a result that could have been antici-

pated in view of the relation between the old and new overlaps. The transfer matrix is strictly positive for $Ma_5<1$. (Thus, for $M<2$ it is sufficient to take e.g., $a_5=0.5$.) In this case, only close-to-unity eigenvalues [4,10,12,16,17] may lead to significant anomalous contributions. We propose a method to reduce the density of such eigenvalues in dynamical simulations. The method consists of a modification of the pseudo-fermion (Pauli-Villars) part of the action, and it might be effective already for modest values of N_s. The main side effect of the modified action is a small renormalization of the coupling constant.

This paper is organized as follows. Domain-wall fermions with a variable lattice spacing for the fifth coordinate are reviewed in Sec. II. The (first quantized) transfer matrix is introduced in Sec. III. The case of a complex transfer matrix is discussed in Sec. IV. The modified pseudo-fermion action is introduced and discussed in Sec. V. Our conclusions are given in Sec. VI, and some technical details are relegated to two appendices.

II. DOMAIN-WALL FERMIONS WITH A VARIABLE *s*-SPACING

Allowing for a variable lattice spacing a_5 for the fifth coordinate [15], the domain-wall fermion matrix is

$$D_F=\begin{pmatrix} a_5D-1 & P_R & 0 & 0 & \dots & 0 & 0 & -ma_5P_L \\ P_L & a_5D-1 & P_R & 0 & \dots & 0 & 0 & 0 \\ 0 & P_L & a_5D-1 & P_R & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & P_L & a_5D-1 & P_R \\ -ma_5P_R & 0 & 0 & 0 & \dots & 0 & P_L & a_5D-1 \end{pmatrix}. \tag{2.1}$$

The four-dimensional lattice spacing a is set to unity. The above matrix structure corresponds to the fifth coordinate s, which we assume to take the values $s=1,2,\dots,N_s$. Each entry is a four-dimensional matrix. $P_{R,L}=\frac{1}{2}(1\pm\gamma_5)$ denote chiral projectors, and D is the Wilson-Dirac operator

$$D=\begin{pmatrix} M-W & C \\ -C^\dagger & M-W \end{pmatrix}, \tag{2.2}$$

where

$$C_{xy}=\frac{1}{2}\sum_\mu [\,\delta_{x+\hat{\mu},y}U_{x\mu}-\delta_{x-\hat{\mu},y}U^\dagger_{y\mu}]\sigma_\mu \tag{2.3}$$

$$W_{xy}=4\delta_{xy}-\frac{1}{2}\sum_\mu [\,\delta_{x+\hat{\mu},y}U_{x\mu}+\delta_{x-\hat{\mu},y}U^\dagger_{y\mu}]. \tag{2.4}$$

The matrices C and W correspond to the kinetic and Wilson term respectively. We also define

$$B_{xy}=(1-Ma_5)\delta_{xy}+a_5W_{xy}. \tag{2.5}$$

(B is proportional to the identity matrix in spinor space. The latter may be either two-by-two or four-by-four; the correct meaning can be inferred from the context.)

Assume momentarily a semi-infinite range for the s-coordinate. For $\sin(p_\mu)=0$, $\mu=1,\dots,4$, the free-field equation has a right-handed homogeneous solution

$$\psi^0_R(s;p_\mu)=B^s_0(p_\mu)=[1+a_5(W_0(p_\mu)-M)]^s, \tag{2.6}$$

where the subscript zero denotes free-field quantities, $W_0(p_\mu)=\Sigma_\mu(1-\cos(p_\mu))$, and at the corners of the Brillouin zone $W_0(p_\mu)$ takes the values $0,2,\dots,8$. The above homogeneous solution is a zero mode (i.e., it is normalizable) provided

$$-1<1+a_5(W_0(p_\mu)-M)<1. \tag{2.7}$$

We depart from the original domain-wall framework by replacing the constraint on the lower bound with the stronger one

$$0<1-Ma_5. \tag{2.8}$$

Since W is a positive matrix, this implies

$$0<1+a_5(\mathrm{spec}(W)-M). \tag{2.9}$$

The last condition ensures the strict *positivity* of B. The upper-bound constraint in Eq. (2.7) is simply

$$W_0(p_\mu)-M<0, \tag{2.10}$$

which is evidently independent of a_5. For $0<M<2$ there is a single zero mode at $p_\mu=0$. If we increase M while decreasing a_5 to maintain the constraint (2.8), new zero modes will appear at $M=2,4,6,8$, while all the zero modes from smaller values of M will remain in the spectrum.

Assuming $0<M<2$, on a finite lattice the right-handed part of the quark field corresponds to the above zero mode, whereas the left-handed part corresponds to another zero mode located near the $s=N_s$ boundary. Both at the tree level and in perturbation theory, for $m=0$ there is a small mixing

between the two chiral modes that vanishes exponentially with N_s [6,8,9,4]. The parameter m in Eq. (2.1) is related to the bare quark mass. An easy way to see this is to invoke the free domain-wall Hamiltonian. For $m=0$, the eigenvalues $E(p_k)$ of helicity eigenstates are given by $E^2=\Sigma_{k=1}^{3}\sin^2(p_k)$. For $m\neq 0$ we find using first-order perturbation theory that $E(0)=m_q$ where

$$m_q = mMa_5(2-Ma_5), \tag{2.11}$$

generalizing the result of Ref. [8].

III. PROPAGATORS AND TRANSFER MATRICES

The familiar relation $D=\gamma_5 D^\dagger \gamma_5$, valid for the Wilson-Dirac operator, generalizes in the case of domain-wall fermions to

$$D_F = \mathcal{R}\gamma_5 D_F^\dagger \mathcal{R}\gamma_5, \tag{3.1}$$

where $\mathcal{R}_{ss'}=\delta_{s,N_s+1-s'}$ [10]. Equation (3.1) implies relations between second-order operators: $D_F^\dagger D_F=(\mathcal{R}\gamma_5 D_F)^2=\mathcal{R}\gamma_5(D_F D_F^\dagger)\mathcal{R}\gamma_5$. We also note that $\mathcal{R}\gamma_5 D_F$ is Hermitian. For definiteness we will focus on the operator $D_F D_F^\dagger$. One has

$$D_F D_F^\dagger = \gamma_5 B^{1/2}\Omega\gamma_5 B^{1/2}, \tag{3.2}$$

where explicitly

$$\Omega=\begin{pmatrix} X_{++} & -1 & 0 & 0 & \dots & 0 & 0 & X_{-+} \\ -1 & Y & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & Y & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & Y & -1 \\ X_{+-} & 0 & 0 & 0 & \dots & 0 & -1 & X_{--} \end{pmatrix} \tag{3.3}$$

and

$$Y = 2 + a_5^2 \gamma_5 B^{-1/2} D D^\dagger \gamma_5 B^{-1/2} \tag{3.4}$$

$$X_{++} = Y + P_L B^{-1}((ma_5)^2 - 1) \tag{3.5}$$

$$X_{--} = Y + P_R B^{-1}((ma_5)^2 - 1) \tag{3.6}$$

$$X_{+-} = X_{-+} = ma_5. \tag{3.7}$$

The $B^{\pm 1/2}$ factors have been introduced for later convenience. Thanks to our insistence on the strict positivity of B, these factors are strictly positive too. An interesting observation is that

$$Y = \tilde{T} + \tilde{T}^{-1}, \tag{3.8}$$

where

$$\tilde{T} = \begin{pmatrix} B^{-1} + a_5^2 B^{-1/2} C C^\dagger B^{-1/2} & a_5 B^{-1/2} C B^{1/2} \\ a_5 B^{1/2} C^\dagger B^{-1/2} & B \end{pmatrix}. \tag{3.9}$$

The basic properties of the *transfer matrix* $\tilde{T}$ follow from

$$\tilde{T} = K^\dagger K, \tag{3.10}$$

where

$$K = \begin{pmatrix} B^{-1/2} & 0 \\ a_5 C^\dagger B^{-1/2} & B^{1/2} \end{pmatrix}, \quad K^\dagger = \begin{pmatrix} B^{-1/2} & a_5 B^{-1/2} C \\ 0 & B^{1/2} \end{pmatrix}, \tag{3.11}$$

which imply that $\tilde{T}$ is strictly positive, bounded, and has $\det\tilde{T}=1$.

The first-quantized transfer matrix usually encountered in the context of domain-wall fermions is $T=KK^\dagger$. Evidently, T and $\tilde{T}$ have the same spectrum, and (up to normalization) the eigenvectors of T are obtained from those of $\tilde{T}$ by multiplication with K. The appearance of $\tilde{T}$ instead of T is due to a technical reason that we explain later.

Equations (3.4) and (3.8) imply that an eigenvalue of $\tilde{T}$ is equal to one *iff* the Hermitian operator $\gamma_5 D$ has a zero mode. The last condition implies $\det(D)=0$, an equation which defines a measure-zero subset of the gauge field configuration space. In the rest of this section we will assume that $\det(D)\neq 0$, and hence that no eigenvalue of $\tilde{T}$ is exactly equal to one.

We now turn to the construction of the domain-wall propagator $G_F=D_F^{-1}$. Using Eq. (3.2) one has

$$G_F = D_F^\dagger \gamma_5 B^{-1/2} G \gamma_5 B^{-1/2}, \tag{3.12}$$

where $G=\Omega^{-1}$. Our task is to find an explicit representation for G. We begin by writing the spectral decomposition

$$\tilde{T} = \sum_i |v_i\rangle \lambda_i \langle v_i|. \tag{3.13}$$

For each eigenvalue, we define $q_i=\min(\lambda_i,\lambda_i^{-1})$. We now construct a new matrix

$$Q = \sum_i |v_i\rangle q_i \langle v_i|, \tag{3.14}$$

with the property $0<\mathrm{spec}(Q)<1$. Next we consider the infinite-size matrix Ω_∞ constructed from the translationally invariant part of Ω, extended to the range $-\infty<s,s'<\infty$. The inverse $G_\infty=\Omega_\infty^{-1}$ is

$$G_\infty(s,s') = \sum_i |v_i\rangle q_i^{|s-s'|} f(q_i)\langle v_i| \tag{3.15}$$

$$= Q^{|s-s'|} f(Q), \tag{3.16}$$

where $f^{-1}(q)=q^{-1}-q>0$. Equations (3.3) and (3.8) imply $\Sigma_s \Omega_\infty(s'',s)G_\infty(s,s')=0$ for $s''\neq s'$, while the correct normalization for $s''=s'$ is ensured by the presence of $f(Q)$. Since it was constructed using Q (and not $\tilde{T}$), $G_\infty(s,s')$ vanishes for large $|s-s'|$.

Returning to the finite-s case, we now have, for $1 \leqslant s,s' \leqslant N_s$,

$$G(s,s') = G_\infty(s,s') + H_{++}(s,s') + H_{--}(s,s') + H_{+-}(s,s') + H_{-+}(s,s'), \tag{3.17}$$

where

$$\begin{aligned} H_{++}(s,s') &= Q^s A_{++} Q^{s'} \\ H_{--}(s,s') &= Q^{N_s+1-s} A_{--} Q^{N_s+1-s'} \\ H_{+-}(s,s') &= Q^s A_{+-} Q^{N_s+1-s'} \\ H_{-+}(s,s') &= Q^{N_s+1-s} A_{-+} Q^{s'}. \end{aligned} \tag{3.18}$$

The four-dimensional matrices $A_{\pm\pm}$ solve a system of linear equations given in Appendix A, and have a convergent $N_s \to \infty$ limit. Hermiticity of G implies $A_{-+} = A_{+-}^\dagger$.

The physical significance of the above construction is the following. In a fixed gauge-field background $\mathcal{U} = \{U_{x\mu}\}$ one has $G_F(s,s';\mathcal{U}) = \langle \psi(s) \overline{\psi}(s') \rangle_\mathcal{U}$. Suppose that we project $\overline{\psi}(s')$, the anti-fermion field on a given s'-slice, onto the (four-dimensional) state $\gamma_5 B^{1/2} v_i$ where v_i is an eigenvector of $\tilde{T}$. Using Eqs. (3.12) and (3.17), each term in the corresponding projection of $\langle \psi(s) \overline{\psi}(s') \rangle_\mathcal{U}$ must have an s'-dependence given by either $\lambda_i^{+s'}$ or $\lambda_i^{-s'}$ (λ_i is the corresponding eigenvalue). The s'-dependence will involve no other eigenvalue of $\tilde{T}$. A similar statement can be made for the ψ-field. To this end, we write $G_F = (D_F^\dagger D_F)^{-1} D_F^\dagger$ and use Eq. (3.1) to relate $(D_F^\dagger D_F)^{-1}$ to $(D_F D_F^\dagger)^{-1}$. This reproduces a key feature of the second-quantized transfer matrix formalism [11,10], but now in the Lagrangian formalism, which is directly related to the manner one simulates fermionic correlation functions.

We end this section with two technical comments. In the second-quantized transfer matrix formalism one can define s-dependent operators via $\hat{a}_{L,R}(s) = \hat{T}^s \hat{a}_{L,R} \hat{T}^{-s}$, where $\hat{T}$ is the second-quantized version of T [defined by $\hat{T} = \exp(\hat{a}^\dagger \log(T) \hat{a})$, where $\hat{a}$ and $\hat{a}^\dagger$ are creation and annihilation anti-commuting operators]. As noted in Ref. [10] [see Eq. (A9) there] using the transfer matrix $\hat{T}$ implies that the operators $\hat{a}_R(s)$ and $\hat{a}_L(s)$ are identified with the Grassmann variables $\psi_R(s)$ and $\psi_L(s-1)$ respectively. In other words, related Grassmann variables and operators do not always have the same s-coordinate. On the other hand, if one uses the second-quantized version of $\tilde{T}$, then related Grassmann variables and operators do have the same s-coordinate in all cases. This explains the appearance of $\tilde{T}$ instead of T in the expressions for the domain-wall propagator.

Our last comment concerns the new overlap formalism. Dropping all terms of order a_5^2 and higher in $\tilde{T}$ one has

$$\tilde{T} \sim \begin{pmatrix} 1 + a_5(M-W) & a_5 C \\ a_5 C^\dagger & 1 - a_5(M-W) \end{pmatrix}. \tag{3.19}$$

If we now take the limit $a_5 \to 0$ while keeping the product $a_5 N_s = L_s$ fixed, we have $\tilde{T}^{N_s}(a_5) \to \exp(\gamma_5 D L_s)$. This reproduces the result of Ref. [15]. Recognizing that $\gamma_5 D$ is the new overlap's Hamiltonian, we see that both the old and the new overlap formulas for QCD can be recovered as suitable limits of domain-wall fermions.

IV. COMPLEX TRANSFER MATRIX

In this section we discuss what happens if $Ma_5 > 1$, starting with the free-field case. Since $[B_0, C_0] = 0$, one can lump together $B_0^{\pm 1/2}$ factors, so that only $B_0^{\pm 1}$ will occur in the equations that define Ω_0 and Y_0 [cf. Eqs. (3.2) and (3.4)]. One has

$$Y_0(p_\mu) = B_0^{-1}(p_\mu) \left[1 + B_0^2(p_\mu) + \sum_\mu \sin^2(p_\mu) \right]. \tag{4.1}$$

Note that the sign of $Y_0(p_\mu)$ is determined by the sign of $B_0(p_\mu)$, and that either $Y_0(p_\mu) > 2$ or $Y_0(p_\mu) < -2$. The eigenvalues $\tilde{T}_0(p_\mu)$ are all real, and have the same sign as $B_0(p_\mu)$ [cf. Eq. (3.8)]. Therefore, the free-field propagator exhibits sign oscillations, but otherwise everything stays pretty much the same as in the $Ma_5 < 1$ case.

The situation is different in the interacting theory, where $[B,C] \neq 0$. When $Ma_5 > 1$, B has both positive and negative eigenvalues, and $B^{1/2}$ has both real and imaginary eigenvalues. Consequently, the transfer matrix $\tilde{T}$ is no longer Hermitian. [Equation (3.10) still holds if the *same* definition of $B^{\pm 1/2}$ is used in the expressions for K and $K^\dagger$, cf. Eq. (3.11), even though $B^{1/2}$ is no longer Hermitian. Of course, $K^\dagger$ does *not* stand for the Hermitian conjugate of K in this case.]

While $\det(D_F)$ is always real due to Eq. (3.1), it is not necessarily positive for $Ma_5 > 1$. In particular $\det(D_F)$ may occasionally vanish. (This is not true for $Ma_5 < 1$ and $m > 0$, see Appendix A.) The propagator exists except on the measure-zero subset defined by $\det(D_F) = 0$, and can be constructed using the same technique as in Sec. III. The matrix Q is now defined as follows. Let y_i denote an (in general complex) eigenvalue of Y. For y_i that does not belong to the closed interval $[-2,2]$ on the real axis, we define q_i to be the solution of $q_i + q_i^{-1} = y_i$ obeying $|q_i| < 1$. For $y_i \in (-2,2)$ the roots obey $|q_i| = 1$ while $q_i \neq q_i^{-1}$. In this case we arbitrarily pick one of the roots. Finally, we disregard gauge-field configurations leading to any $y_i = \pm 2$, hence $f(q_i)$ exists. (This amounts to ignoring another measure-zero set; see also the last paragraph of Appendix A.) The rest of the construction is the same as before.

Again, as discussed in Sec. III, the dependence of the propagator on s and s' is governed by the (now complex) eigenvalues and eigenvectors of the transfer matrix $\tilde{T}$. Hence, slowly decreasing anomalous effects may now arise from eigenvalues lying anywhere close to the unit circle.

When $Ma_5>1$, both $B^{-1/2}$ and $\tilde{T}$ are unbounded. The possibility that $\tilde{T}$ may have very large eigenvalues is, however, not worrisome. The latter correspond to very small eigenvalues of Q, and so they lead to very short-range correlations in the s-direction.

Of special significance is the chiral Ward identity that governs the pion mass. For any a_5 and M, the anomalous term in the pion-mass Ward identity is positive when the number of equal-mass dynamical flavors is even, as well as in the quenched case (see Appendix B of Ref. [10]). Thus, the anomalous contributions to this particular Ward identity, coming from all (real or complex) eigenvalues of the transfer matrix, always add up.

Because of Eqs. (3.4) and (3.8), close-to-unity eigenvalues of the non-Hermitian $\tilde{T}$ should still correspond to approximate zero modes of the Hermitian operator $\gamma_5 D$. The latter have been extensively studied recently [4,16,3]. To date, however, no information exists on the distribution of eigenvalue in all the rest of the complex plane. In particular, it is not known how many eigenvalues are located near the unit circle *away* from the point one on the positive real axis. Therefore it is also not known how much of the anomalous effect observed in numerical simulations [1–4] is due to (approximate) zero modes of $\gamma_5 D$. The above issues clearly deserve a more detailed study.

V. A MODIFIED PSEUDO-FERMION ACTION

In the rest of this paper we impose the condition $Ma_5<1$, and alongside with it the strict positivity of the transfer matrix. For simplicity we also restrict the discussion to $0<M<2$. As discussed in the Introduction, the troublesome eigenvalues of the transfer matrix are now the ones close to unity. In this section we propose a method to reduce the density of close-to-unity eigenvalues in a dynamical simulation by modifying the *pseudo-fermion* (also known as Pauli-Villars) part of the action.

Let us first recall why a domain-wall fermion action must be accompanied by a pseudo-fermion action. From the point of view of the gauge field, the domain-wall action introduce N_s ''flavors'' of four-dimensional Dirac fermions. For $0<M<2$, only one Dirac fermion is light (and is identified with a quark field). The other N_s-1 Dirac fields have $O(1)$ masses. If their number was kept fixed, we could simply ignore them in the continuum limit. The chiral limit, however, requires $N_s\to\infty$. If this limit is taken at fixed value of the bare coupling g, the N_s-dependent contribution of the heavy ''flavors'' must be subtracted.

One can express the domain-wall fermion determinant as

$$\det(D_F)=\mu_F^{N_s}\times(\text{finite factor}). \tag{5.1}$$

The ''finite factor,'' which accounts for the quark field, has a convergent $N_s\to\infty$ limit. The bulk term $\mu_F^{N_s}$ is the (undesirable) contribution of the $O(N_s)$ massive ''flavors.'' Explicit expressions for both terms were first derived in the transfer matrix formulation [11,10], and more recently by direct manipulations of determinants [18]. The latter technique will also be used below. Explicitly,

$$\mu_F=\prod_{\lambda_i>1}\lambda_i, \tag{5.2}$$

where the product is over the greater-than-one eigenvalues of $\tilde{T}$, cf. Eq. (3.9). As expected, the bulk term is independent of the choice of boundary conditions in the s-direction (i.e., it is independent of m). One way to cancel the bulk contribution is to introduce a five-dimensional boson field ϕ_{xs}, having the same spin and internal indices as the domain-wall fermions, on a lattice whose fifth coordinate ranges only from 1 to $N=N_s/2$. We will refer to ϕ_{xs} as the pseudo-fermion field. The pseudo-fermion action is

$$S_{\rm pf}^{\rm unmodified}=\sum_{xs,ys'}\phi^\dagger_{xs}(D_{\rm pf}D^\dagger_{\rm pf})_{xs,ys'}\phi_{ys'}. \tag{5.3}$$

A common choice is $D_{\rm pf}=D_F(ma_5=1)$. Excepting $ma_5=-1$, in fact, any $O(1)$ value for ma_5 will do. When integrating over both fermions and pseudo-fermions the bulk terms cancel out, leaving a convergent $N_s\to\infty$ result that has the same long-distance behavior in four dimensions as the finite factor in Eq. (5.1). (Another option is to use a first-order pseudo-fermion action. When the transfer matrix has a gap, the bulk term of the first-order action converges faster to the fermionic bulk term [4]. This advantage disappears if the gap is small.)

As mentioned above, for $Ma_5<1$ chiral symmetry violations should arise from close-to-unity eigenvalues of the transfer matrix. Actually, in the free-field case the spectrum has a gap. This follows from Eqs. (3.4) and (3.8), and the fact that

$$D_0D_0^\dagger(p_\mu)=(M-W_0(p_\mu))^2+\sum_\mu\sin^2(p_\mu)>0. \tag{5.4}$$

Since the Brillouin zone is compact, the (M-dependent) minimal eigenvalue of $D_0D_0^\dagger$ is strictly positive. Now, the eigenvalues of $DD^\dagger$ are gauge-invariant continuous functions of the link variables. Since $D_0D_0^\dagger$ has a gap in the free-field case, one expects that $DD^\dagger$ too should have a (somewhat smaller) gap if the local plaquette action

$$\mathcal{P}_x=\sum_{\mu<\nu}\mathrm{Re\,tr}(1-U_{x\mu}U_{x+\hat{\mu},\nu}U^\dagger_{x+\hat{\nu},\mu}U^\dagger_{x\nu}), \tag{5.5}$$

is very small everywhere. In other words, zero modes of $DD^\dagger$ should arise only if $\mathcal{P}_x$ exceeds an $O(1)$ constant $c_0(M)>0$ at least for *some* lattice sites. Hence, if one is sufficiently close to the continuum limit and M is not too close to the critical points 0 or 2, these zero modes should be suppressed by the plaquette action. In the context of domain-wall fermions we can therefore regard these zero modes as lattice artefacts. [See Refs. [16,17] for related work. The existence of a $c_0(M)>0$ can be tested numerically. A proof

that $c_0(M)>0$ is desirable, but one should keep in mind that rigorous inequalities may underestimate the value of $c_0(M)$.]

For $\beta\sim 6$, one is probably not close enough to the continuum limit to effectively suppress the zero modes of $DD^\dagger$ by the plaquette action alone. The above consideration leads us to propose a modified pseudo-fermion action

$$S_{\rm pf}=\sum_{xs,ys'}\phi^\dagger_{xs}(D_{\rm pf}D^\dagger_{\rm pf})_{xs,ys'}\phi_{ys'}+\sum_{xs}\phi^\dagger_{xs}(c_1\mathcal{P}_x)^n\phi_{xs}, \tag{5.6}$$

where c_1 is a continuous parameter and n is a positive integer. Assuming $c_0(M)$ is known, a reasonable choice is $c_1\sim c_0^{-1}(M)$. The choice of n is discussed later. The idea behind the modified pseudo-fermion action is to suppress in a selective way those gauge-field configurations supporting close-to-unity eigenvalues. The aim is to achieve this suppression in (dynamical-fermion) simulations with not too large N_s and at presently accessible values of β. As explained below, the main side effect of the modified action is expected to be a small renormalization of the coupling constant.

As in Eq. (3.2), the pseudo-fermion matrix defined by Eq. (5.6) can be written as $\gamma_5 B^{1/2}\Omega_{\rm pf}\gamma_5 B^{1/2}$. The explicit expression for $\Omega_{\rm pf}$ is the same as the right-hand side (RHS) of Eq. (3.3) except for the replacement

$$Y\to Y_{\rm pf}(c_1)=Y+B^{-1/2}(c_1\mathcal{P})^nB^{-1/2}, \tag{5.7}$$

where $\mathcal{P}$ stands for the diagonal matrix $\delta_{xy}\mathcal{P}_x$. (Similar replacements are made in the definitions of X_{++} and X_{--}. For $ma_5=1$ one has $Y_{\rm pf}=X^{\rm pf}_{++}=X^{\rm pf}_{--}$.)

An expression for $\det(\Omega_{\rm pf})$ can be written for even N using the general formulas of Ref. [18], see Appendix B below. The result is

$$\det(\Omega_{\rm pf})=\det(R^- - PT^N_{\rm pf}P^{-1}R^+), \tag{5.8}$$

where

$$R^-=\begin{pmatrix}1 & 0\\ Y_{\rm pf}-X^{\rm pf}_{--} & -X_{+-}\end{pmatrix} \tag{5.9}$$

$$R^+=\begin{pmatrix}-X_{-+} & Y_{\rm pf}-X^{\rm pf}_{++}\\ 0 & 1\end{pmatrix} \tag{5.10}$$

$$T_{\rm pf}=\begin{pmatrix}Q_{\rm pf} & 0\\ 0 & Q^{-1}_{\rm pf}\end{pmatrix} \tag{5.11}$$

$$P=\begin{pmatrix}1 & Q_{\rm pf}\\ Q_{\rm pf} & 1\end{pmatrix},\quad P^{-1}=(1-Q^2_{\rm pf})^{-1}\begin{pmatrix}1 & -Q_{\rm pf}\\ -Q_{\rm pf} & 1\end{pmatrix}. \tag{5.12}$$

As before, the matrix $Q_{\rm pf}$ is defined by $Q_{\rm pf}+Q^{-1}_{\rm pf}=Y_{\rm pf}$ and the condition $0<\mathrm{spec}(Q_{\rm pf})<1$. One has

$$PT^2_{\rm pf}P^{-1}=\begin{pmatrix}-1 & Y_{\rm pf}\\ -Y_{\rm pf} & Y^2_{\rm pf}-1\end{pmatrix}. \tag{5.13}$$

Note that the RHS of the last equation is a function of $Y_{\rm pf}$ only.

For large N, $Q^N_{\rm pf}$ vanishes whereas $Q^{-N}_{\rm pf}$ grows unboundedly. Like Eq. (5.1), one can express $\det(\Omega_{\rm pf})$ as $\mu^N_{\rm pf}(c_1)$ times a finite factor, where

$$\mu_{\rm pf}(c_1)=\det(Q^{-1}_{\rm pf}(c_1)). \tag{5.14}$$

[See Eq. (B7) for the finite factor.] For $N_s=2N$, the total bulk factor coming from the integration over both fermions and pseudo-fermions is

$$\left(\frac{\mu^2_F}{\mu_{\rm pf}(c_1)}\right)^N. \tag{5.15}$$

In the unmodified case, $c_1=0$, using Eq. (5.2) and $\det(\tilde T)=1$ one has $\mu_{\rm pf}(0)=\mu^2_F$, showing that the bulk factors indeed cancel each other.

We now discuss how the bulk factor is modified for $c_1>0$. Since the plaquette term is positive, the c_1-derivative of the eigenvalues of $Y_{\rm pf}$ is always positive, and the c_1-derivative of the eigenvalues of $Q_{\rm pf}$ is always negative. Therefore $\mu^2_F/\mu_{\rm pf}(c_1)$ is a decreasing function of c_1.

Let us now assume that $\tilde T$ (or Q) has an eigenvalue very close to one. For $M\sim 1.7$, the support of the corresponding eigenvector should consist of very few lattice sites [16]. Moreover, as discussed earlier $\mathcal{P}_x$ is likely $O(1)$ on those sites. As c_1 is varied from zero to its chosen value, we thus expect an $O(1)$ change in the corresponding eigenvalue of $Q_{\rm pf}(c_1)$. We will argue below that there should be only a small change in the product of all eigenvalues which are *not* close to unity. Consequently, the $O(1)$ change in what used to be a close-to-unity eigenvalue implies an $O(1)$ reduction in the magnitude of $\mu^2_F/\mu_{\rm pf}(c_1)$. This, in turn, should suppress the Boltzmann weight of the corresponding gauge-field configuration already for modest values of N_s [cf. Eq. (5.15)].

Eigenvalues of the transfer matrix not too close to unity are typically not localized, and it is plausible that the effect of the modified action on them can be accounted for by perturbation theory. It is easy to see that the leading perturbative effect of the modified action is to renormalize the coupling constant as

$$\frac{1}{g^2}\to\frac{1}{g^2}+nc_1^n\sum_s\langle\mathcal{P}^{n-1}_x\phi^\dagger_{xs}\phi_{xs}\rangle. \tag{5.16}$$

Since the free transfer matrix has a gap, the perturbative value of $\langle\phi^\dagger_{xs}\phi_{xs}\rangle$ at any finite order, and in any gauge-field background, is regular. (This is true even if the associated exact transfer matrix has a unit eigenvalue. In other words, a singularity in the propagator cannot develop if we sum the Born series only up to a finite order.) Now, the expectation value on the RHS of Eq. (5.16) involves n or more loops. Therefore the resulting change in $1/g^2$ should be of order

$N_s\alpha^n$. Since in practice $\alpha\sim 0.1$, if we take for example $n=4$ this effect may be at the level of 1% or less, when N_s is in the range of 10 to 100. Thus, the leading side-effect of the modified action seems to be an innocuous, relatively small change in the bare coupling.

In short, we believe that perturbation theory can be trusted for the collective contribution of all eigenvalues, *except* when there are close-to-unity ones. Since the perturbative effect should be small, an $O(1)$ change in $\mu_{\rm pf}(c_1)$ should take place only when there are close-to-unity eigenvalues, and this change works in the direction of suppressing the Boltzmann weight of the corresponding gauge-field configurations.

The above arguments, while plausible, are heuristic. One question that can be settled by an explicit (perturbative) calculation is whether the loop integrals in $\langle \mathcal{P}_x^{n-1}\phi_{xs}^\dagger\phi_{xs}\rangle$, while regular, happen to produce large numerical factors. A calculation of this expectation value is also necessary in order to be able to compare results at different values of c_1, while maintaining a fixed value of the effective bare coupling [cf. the RHS of Eq. (5.16)].

Last we discuss how the $N_s\to\infty$ limit may be taken with the modified action. Evidently, if all other parameter were kept fixed, then in the limit $N_s\to\infty$ the perturbative effects induced by the modified pseudo-fermion action would eventually run out of control. While the strict $N_s\to\infty$ limit is not very useful for practical purposes, it is legitimate to ask whether, in principle, the modified action has a sensible $N_s\to\infty$ limit.

If we allow both N_s and n to grow, the limit may in fact depend on the ratio of these two numbers. As an example, one option is to take the limit $n\to\infty$ *before* the limit $N_s\to\infty$. Sending n to infinity has the following effect. For $c_1\mathcal{P}_x<1$, the limit of $(c_1\mathcal{P}_x)^n$ is zero. Hence, the modification vanishes if $c_1\mathcal{P}_x<1$ for all x. If, on the other hand, $c_1\mathcal{P}_x>1$ even for a single site, the norm of $(c_1\mathcal{P}_x)^n$ will blow up, and along with it some of the eigenvalues of $Q_{\rm pf}^{-1}$. In summary, in the limit $n\to\infty$ the bulk factor $\mu_F^2/\mu_{\rm pf}(c_1)$ is unchanged if $c_1\mathcal{P}_x<1$ for all x, whereas it vanishes if $c_1\mathcal{P}_x>1$ for any lattice site. Therefore, the limit $n\to\infty$ amounts to imposing the constraint $c_1\mathcal{P}_x<1$ on the gauge-field configuration space. While this constraint should not change the continuum limit, throwing out all configurations with some $c_1\mathcal{P}_x>1$ in a completely unselective manner could slow down simulations. As argued above, for moderate values of N_s and relatively small values of n, the modified action may do a good job in suppressing only those gauge-field configurations supporting close-to-unity eigenvalues, leading to a minimal ''waste'' of configurations.

VI. CONCLUSIONS

In this paper we have studied potential sources for the anomalous term in chiral Ward identities. For conventional domain-wall fermions ($a_5=1$) the transfer matrix is complex if $M>1$. As discussed in Sec. IV, in the numerical work of Refs. [1–4] slowly-suppressed chiral symmetry violations could therefore arise not only from eigenvalues which are close to one, but in fact from eigenvalues in the vicinity of the entire unit circle.

For $a_5=0.5$ and $0<M<2$ the transfer matrix will be strictly positive. The first question that has to be addressed numerically is whether the favorable range for a single light quark is still $M\sim 1.7$, as in the $a_5=1$ case [1,3]. The range $M\sim 1.7$ seems reasonable also from the point of view of the new overlap formulation [16]. Since the new overlap involves a continuous s-coordinate limit, it is plausible that the best value of M may be rather insensitive to a_5.

For fixed $1<M<2$, as a_5 is decreased from 1 to 0.5 all eigenvalues must flow towards the positive real axis. Depending on the flow pattern, the number of close-to-one eigenvalues at $a_5=0.5$ could be quite different from the original number of eigenvalues close to the unit circle. Hence, the transition from $a_5=1$ to $a_5=0.5$ could by itself have a significant effect on the anomalous term. Finally, once the above issues are resolved, one can proceed to test whether the modified pseudo-fermion action is useful in further reducing lattice-artifact violations of chiral symmetries at presently accessible values of N_s.

ACKNOWLEDGMENTS

I thank the participants of the RIKEN BNL Research Center workshop on ''Fermion Frontiers in Vector Lattice Gauge Theories,'' at Brookhaven National Laboratory, for extensive discussions that motivated this work. Especially I would like to thank the organizers for creating this unique opportunity to exchange ideas on the subject. This work is supported in part by the Israel Science Foundation.

APPENDIX A: THE SECOND-ORDER PROPAGATOR

For $2\leqslant s''\leqslant N_s-1$, Eqs. (3.3) and (3.8) imply

$$\sum_s \Omega(s'',s)G_\infty(s,s')-\delta_{s'',s'}=\sum_s \Omega(s'',s)H_{\pm\pm}(s,s')=0. \tag{A1}$$

For $s''=1$ and $s''=N_s$ there are boundary effects. The above expressions are not zero, and their s'-dependence is given by $Q^{\pm s'}$. The linear combination that gives the propagator G [Eq. (3.17)] is determined by requiring the coefficients of $Q^{\pm s'}$ to vanish. By imposing this condition on the $Q^{+s'}$ part we obtain

$$\mathcal{C}\begin{pmatrix}A_{++}\\ A_{-+}\end{pmatrix}=\begin{pmatrix}Q^{-2}-X_{++}Q^{-1}\\ -X_{+-}Q^{-1}\end{pmatrix}f(Q), \tag{A2}$$

and by imposing this condition on the $Q^{-s'}$ part we obtain

$$\mathcal{C}\begin{pmatrix}A_{+-}\\ A_{--}\end{pmatrix}=\begin{pmatrix}-X_{-+}Q^{-1}\\ Q^{-2}-X_{--}Q^{-1}\end{pmatrix}f(Q), \tag{A3}$$

where

$$\mathcal{C}=\begin{pmatrix} X_{++}Q-Q^2+X_{-+}Q^{N_s} & X_{++}Q^{N_s}-Q^{N_s-1}+X_{-+}Q \\ X_{--}Q^{N_s}-Q^{N_s-1}+X_{+-}Q & X_{--}Q-Q^2+X_{+-}Q^{N_s} \end{pmatrix}. \tag{A4}$$

Note the $\mathcal{C}$ has a convergent $N_s\to\infty$ limit. A second-quantized transfer-matrix representation of the finite-N_s fermion determinant is given in Ref. [10] [see in particular Eqs. (3.1) and (3.10) therein]. We conclude from it that for $Ma_5<1$ and $m>0$, $\det(D_F)$ is (real and) strictly positive for any number of flavors. Therefore D_F has an inverse, and the above equations must have a solution which obeys $A_{-+}=A_{+-}^{\dagger}$.

We note that D_F^{-1} may exist even if $\tilde{T}$ has an eigenvalue which is exactly equal to one. In this case, a possible way to construct D_F^{-1} is to perturb the background field so that that eigenvalue will be only approximately equal to one, say up to $O(\epsilon)$. The construction of Sec. III is now applicable. D_F^{-1} for the initial background field can then be found by carefully removing the perturbation, keeping track of the leading terms in ϵ.

APPENDIX B: THE SECOND-ORDER DETERMINANT

For even N, we can use the formulas for the determinant of a general tridiagonal matrix, derived in the appendix of Ref. [18], to write down an expression for $\det(\Omega_{\rm pf}(c_1))$, cf. Eq. (5.8). [Here the block entries of the tridiagonal matrix have spin indices ranging from one to four. With the replacement $N\to N_s$, a similar formula holds for $\det(\Omega)$, cf. Eq. (3.3).] In the notation of Ref. [18], the present application is defined by

$$\alpha=\begin{pmatrix} -1 & 0 \\ Y_{\rm pf} & -1 \end{pmatrix}=\alpha_j, \quad j=1,\ldots,N/2-1, \tag{B1}$$

$$\alpha_{N/2}=\begin{pmatrix} -1 & 0 \\ X^{\rm pf}_{--} & X^{\rm pf}_{+-} \end{pmatrix}, \tag{B2}$$

$$\beta=\begin{pmatrix} -1 & Y_{\rm pf} \\ 0 & -1 \end{pmatrix}=\beta_j, \quad j=1,\ldots,N/2-1, \tag{B3}$$

$$\beta_{N/2}=\begin{pmatrix} X^{\rm pf}_{-+} & X^{\rm pf}_{++} \\ 0 & -1 \end{pmatrix}. \tag{B4}$$

With these definitions, Eq. (A.10) of Ref. [18] reads

$$\det(\Omega_{\rm pf})=\det[\alpha^{-1}\alpha_{N/2}-(-\alpha^{-1}\beta)^{N/2}\beta^{-1}\beta_{N/2}]. \tag{B5}$$

Substituting the above explicit expressions we arrive at Eq. (5.8). Note that $(-\alpha^{-1}\beta)$ is equal to the RHS of Eq. (5.13). The bulk factor of $\det(\Omega_{\rm pf})$ can be separated as follows. We first rewrite Eq. (5.8) as

$$\det(\Omega_{\rm pf})=\det(1-Q_{\rm pf}^2)\det(S^- - T_{\rm pf}^N S^+), \tag{B6}$$

where $S^{\pm}=P^{-1}R^{\pm}$. Using Eq. (5.11) it is now easy to check that

$$\det(\Omega_{\rm pf})=\det(Q_{\rm pf}^{-N})\det(1-Q_{\rm pf}^2)\det\left[\begin{pmatrix} 1 & 0 \\ 0 & Q_{\rm pf}^N \end{pmatrix}S^- - \begin{pmatrix} Q_{\rm pf}^N & 0 \\ 0 & 1 \end{pmatrix}S^+\right]. \tag{B7}$$

The bulk factor $\det(Q_{\rm pf}^{-N})$ appears explicitly in the above equation. The other terms are by definition the finite factor. For $Ma_5<1$ the entries of $\Omega_{\rm pf}$ are bounded, and $\det(\Omega_{\rm pf})$ is finite for any finite N. This is more easily seen using Eqs. (5.8) and (5.13). It can be checked that if an eigenvalue of Q approaches one, then the restriction of the RHS of Eq. (5.13) to this eigenvalue has a finite limit. Moreover, the restriction of $PT_{\rm pf}^N P^{-1}$ to this eigenvalue grows only linearly (not exponentially) with N.

[1] T. Blum and A. Soni, Phys. Rev. D **56**, 174 (1997); Phys. Rev. Lett. **79**, 3595 (1997).
[2] T. Blum, A. Soni, and M. Wingate, presented at Lattice' 98.
[3] G.R. Fleming, presented at Lattice '98; R.D. Mawhinney, *ibid.*
[4] P. Vranas, Phys. Rev. D **57**, 1415 (1998); presented at Lattice '98.
[5] D. B. Kaplan, Phys. Lett. B **288**, 342 (1992).
[6] K. Jansen, Phys. Rep. **288**, 348 (1992); Phys. Lett. B **273**, 1 (1996).
[7] R. Narayanan and H. Neuberger, Phys. Lett. B **302**, 62 (1993).
[8] Y. Shamir, Nucl. Phys. **B406**, 90 (1993).
[9] S. Aoki and Y. Taniguchi, Phys. Rev. D (to be published), hep-lat/9711004; Y. Kikukawa, H. Neuberger, and A. Yamada, Nucl. Phys. **B526**, 572 (1998).
[10] V. Furman and Y. Shamir, Nucl. Phys. **B439**, 54 (1995).
[11] R. Narayanan and H. Neuberger, Nucl. Phys. **B412**, 574 (1994).
[12] R. Narayanan and H. Neuberger, Nucl. Phys. **B443**, 305 (1995).
[13] H. Neuberger, Phys. Lett. B **417**, 141 (1998).
[14] R.G. Edwards, U.M. Heller, and R. Narayanan, Nucl. Phys. **B522**, 285 (1998).
[15] Y. Kikukawa and A. Yamada, hep-lat/9806013.
[16] R.G. Edwards, U.M. Heller, and R. Narayanan, hep-lat/9802016.
[17] S. Sharpe and R. Singleton, Phys. Rev. D **58**, 074501 (1998).
[18] H. Neuberger, Phys. Rev. D **57**, 5417 (1998).